DOCUMENT RESUME

ED 194 641 TM 800 779

AUTHOR	De Lisi, Richard
TITLE	Intelligence, Intelligence Testing, and School Practices. ERIC/TM Report 74.
INSTITUTION	ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	ERIC-TM-74
PUB DATE	Dec 80
CONTRACT	400-78-0003
NOTE	51p.
AVAILABLE FROM	ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, Princeton, N.J. 08541 ($4.50).
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Ability Grouping: *Educational Practices: Elementary Secondary Education: *Intelligence: *Intelligence Quotient: *Intelligence Tests: Psychological Evaluation: Scores: Test Bias: Test Format: *Testing: *Test Interpretation: Test Results: Test Validity: Track System (Education)
IDENTIFIERS	*Binet Simon Intelligence Scale

ABSTRACT

A review of the basic information on intelligence testing as a schooling practice is presented, including a discussion of what intelligence is and three psychological approaches to studying intelligence (psychometric, behavioristic, and developmental). All intelligence tests derived from the psychometric approach are standardized and norm-referenced. Some properties that tests do not have in common and which result in different types of intelligence tests include: (1) individually administered tests: (2) group-administered tests: (3) global versus specialized measures: and (4) measures of ability versus achievement. The distinction between aptitude and achievement is a source of controversy in the field of educational measurement that is far from being resolved. In terms of schooling practices, the inability to distinguish between the ability and achievement components of intelligence test performance is important to recognize. A "below average" score on an intelligence test may reflect a lack of opportunity for a child to learn what is required on the test, or it may reflect a slow rate of picking up information from the environment due to learning difficulties of the child. The former problem is one of achievement: the latter is one of ability. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED194641

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ERIC/TM Report 74

INTELLIGENCE, INTELLIGENCE TESTING, AND SCHOOL PRACTICES

Richard De Lisi

Rutgers University

ERIC CLEARINGHOUSE ON TESTS, MEASUREMENT, & EVALUATION

EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY 08541

TM 800 779

The material in this publication was prepared pursuant to a contract with the National Institute of Education, U.S. Department of Education. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to qualified professionals for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either these reviewers or the National Institute of Education.

CONTENTS

Preface..... i

Introduction..... 1

What is intelligence?..... 3

Studying intelligence: Three psychological approaches..... 6

Testing as a schooling practice..... 10

Binet's seminal work..... 14

Properties and types of intelligence tests..... 17

Intelligence test scores: What is IQ?..... 25

Individual and group differences in IQ..... 27

Do IQ tests really measure intelligence?..... 30

Are IQ tests biased?..... 32

Alternatives to IQ testing..... 34

References..... 40

Appendix..... 42

PREFACE

This paper is intended for parents and teachers who want to know more about intelligence testing as a schooling practice. My training as a developmental psychologist has taught me that in order to understand something we must be able to describe and explain its origins and subsequent development. I have approached the subject of intelligence testing with this perspective in mind. The focus is not so much on historical facts but rather on an account of the factors that led to intelligence testing and its use in schools. In my opinion, if you know why and how intelligence tests were first constructed, you will be in a better position to understand their strengths and limitations.

Having said a little about what this paper includes, let me indicate what it omits The paper is not intended to be a comprehensive parent's guide to intelligence testing. A complete inventory of types of tests which details their compositions and scoring systems is not reported here. Suggestions for ways to improve your IQ or your child's IQ are also not included in these pages. There are numerous books on the market that address these topics. Unfortunately, they do not always help the reader understand what the tests are in terms of the functions they serve in schools. This paper explicitly focuses on intelligence testing as a schooling practice. The use of the tests as a clinical tool is not covered, but useful references on this subject are provided. I would like to thank Jeffrey K. Smith and two anonymous reviewers for their critical comments and helpful suggestions which served to improve this paper.

Richard De Lisi

Rutgers University

INTRODUCTION

The majority of adults born in the United States after 1920 have taken at least one, if not several, intelligence tests during their lives. Until recently, intelligence tests were routinely administered as screening and placement devices for educational and occupational programs. During the past few years, such use of tests has been challenged in our legal system. In some states, such as California, IQ-based classifications and program placements of children in schools have been outlawed. In other states, such as Illinois, the courts have ruled "in favor" of IQ tests. That is, a judge has decided that schools can administer intelligence tests and use the results for special education program placement.

These conflicting legal decisions make it apparent that our society is in a state of flux regarding the use of intelligence tests. Because court cases such as Larry P., Bakke, and PASE have been publicized in the news media, this state of affairs is widely recognized. Many people find themselves taking a stand on the use of tests without a good grasp of the issues that surround the current controversy. This lack of understanding has not led to a calm search for the facts. Instead, the topic is rather inflammatory. For some, the very term IQ is synonymous with elitism and racism. For others, IQ represents one of psychology's most successful scientific accomplishments. Even experts in the field of educational testing and measurement cannot agree on some of the "IQ controversies." Because scientists disagree on important points about intelligence tests, the courts will decide their fate (Scarr 1978).

The purpose of this paper is to provide basic information on intelligence testing as a schooling practice. If the paper helps to dispel some of the prevailing misconceptions and thereby make for a more rational process of

decision making, it will have been a success. Toward that end, the first thing for the reader to realize is that there are few absolute "rights or wrongs" on the subject of intelligence testing. For example, a case can be made that the tests are culturally biased. On the other hand, the opposite case that the tests are not biased can also be made. In fact, these conflicting conclusions were reached by the California and Illinois judges, respectively.

Consider the following questions:

What is intelligence?

Are there innate differences in intellectual ability?

Are intelligence tests culturally biased?

Should schools use intelligence tests for program placement?

It seems that answers to such questions depend as much on political and philosophical views of child development as they do on knowledge of the facts. More precisely, different interpretations are made from existing research findings by the experts due to differences in political ideology (Gage and Berliner 1979).

Decisions about the use of intelligence tests, then, are not just a matter of getting acquainted with terms, facts, and concepts from educational psychology. Instead, these decisions are largely a matter of one's views of society's obligations to provide a real chance for its children in the schools. Should children be made to adapt to schools or should schools adapt to the widely disparate abilities of individual children? If children must fit the academic mold, when does it make sense to evaluate and compare children and to use these results for program placement? If schools are to adapt to differences among children, is our society willing to endure the struggles and costs to meet this challenge?

It is hoped that upon completion of this paper the reader will see the relevance of these questions for the future of intelligence testing and be in a better position to answer them intelligently.

WHAT IS INTELLIGENCE?

It is difficult to define intelligence in a simple, straightforward manner. All of the following, either alone or in combination, fit our conception of what intelligence is:

- the ability to learn
- the ability to adapt to new situations
- the ability to perceive relationships and solve problems
- the ability to deal with abstractions and symbols

Despite this lack of definitional consensus, most of us feel we know what intelligence is and can recognize it when we see it. Also, we tend to think of intelligence as something that is innately determined and relatively fixed. The following exchange between two adults captures this everyday view of intelligence. "Mr. Jones, I've been watching your boy for some time now. I've noticed that he is really very bright for his age." "Of course, Mrs. Smith. He's my boy, isn't he?"

Although social scientists have not agreed upon one definition of intelligence, they do agree on the following three points.

(1) Cultural values. Whatever one's criterion for or definition of intelligence is, it is partly a function of the values of the given culture. Since values vary from one culture to the next, so do definitions of intelligence. What this means is that although most cultures have a concept of "intelligent

persons," that concept differs across cultures. An intelligent person in one culture may not be an intelligent person in another culture (Goodnow, in press; Horn 1979).

All of the definitions of intelligence listed above are consistent with and reflect the values of twentieth-century Western civilization. Our culture views abstract thinking as more intelligent than concrete thinking, especially as reflected in the use of words, numbers, and other symbols. We think of intelligence as the ability to learn in a symbolic rather than concrete medium (Gage and Berliner 1979).

(2) Ability-performance distinction. In the study of traits such as intelligence, a distinction is made between "the ability" and "manifestations of that ability." When we judge that a person is intelligent, it is because we see that person perform or behave in a manner that is consistent with our definition of intelligence. We observe performance and then make an inference about ability. It is difficult to prove conclusively that one individual has more ability than another individual. It is easy to demonstrate that there are individual differences in the performance of tasks that are said to require intelligence. Performance is rarely determined by ability alone. The person with "greater ability" (to play a musical instrument, run a race, or think abstractly) will not always outperform counterparts with "lesser ability" due to factors such as fatigue, motivation, familiarity with the task, and so on.

(3) Evidence for validity. This point about ability and performance is more than a matter of semantics. It has implications for how social scientists measure and study human traits. In order to check that performance in one situation (such as on a test) is an accurate reflection or true measure

of ability, scientists examine performance in a different situation to see whether there is consistency across situations.

For example, suppose an anthropologist developed criteria to measure aggression in a given culture. She might observe several people and score some as more aggressive than others based on her test. But how does the anthropologist know whether one person is truly more aggressive than another? Said another way, how does she know that her test of aggression is valid? One way to check is to take people who differed on the test and to observe them in another setting, such as during leisurely activity. If those scored as more aggressive on the anthropologist's test are found to be more aggressive in their spare-time activities, she would have evidence that her test was a valid measure of an ability or trait called aggression. Note that the anthropologist's criteria for aggressive behavior do not have to agree with those of the culture she is studying. Her labels of aggressive or non-aggressive may not match those of the culture itself.

Those who sought to develop tests of intelligence proceeded in a similar fashion. They started with the assumption that individuals differ in the degree to which they have an ability referred to as intelligence. Intelligence was thought to be an ability (or trait) that persons had more or less of in comparison with one another. The ability was thought to be fixed and relatively unchangeable over a person's lifetime. Of course, manifestations of the ability would change with age as the person matured.

Based on these assumptions, psychologists developed tests to measure intelligence. Test items that differentiated among persons of the same and different chronological ages were identified. But how could it be

checked that the items were valid measures of intelligence? It had to be shown that those who were "more intelligent" on the test, i.e., those who could correctly answer more test questions, were also more intelligent in a different setting recognized by psychologists as requiring intelligence. As we shall see later on, one of the more important "other settings" was the school classroom. The fact that people who did well on the newly developed intelligence test items also did well in school (and vice versa for those who did poorly) was taken as evidence that the test was a valid measure of intelligence.

What is the current status of these tests that are designed to measure an ability called intelligence? The situation is somewhat anomalous. On the one hand, psychologists no longer adhere to many of the assumptions about intelligence that their predecessors who made the tests did. For example, the view that intelligence is a fixed, unchangeable trait has been discredited. On the other hand, the evidence for the validity of the tests is strong and impressive. Individuals who differ in performance on intelligence tests are likely to differ in performance in many settings that require intelligence, in a manner that is consistent with the test results. In a word, we may not adhere to some of the beliefs that led to intelligence tests, but we must admit that there is evidence that the tests are valid. In subsequent sections of this paper these points are more fully explained. The next part of the paper reviews three approaches that psychologists have taken to the study of intelligence.

STUDYING INTELLIGENCE: THREE PSYCHOLOGICAL APPROACHES

Once part of philosophy, psychology became a separate field of scientific inquiry during the latter half of the nineteenth century. One of the questions

that the new discipline addressed was the nature of human intelligence. This question came from the older field as philosophers had debated questions of epistemology for centuries. Above all else, psychologists were striving for the attainment of the kind of objectivity that was found in the "mature" sciences such as physics. In particular, psychologists wanted to divorce themselves from the speculative methods and dogma of philosophy. Three distinct "scientific" approaches to the study of intelligence were initiated by psychologists in the late 1890s and early 1900s.

(1) Psychometric approach. This approach to the study of intelligence was part of a larger movement concerned with specification of individual differences in physical and mental characteristics. Darwin's ideas were influential in starting this approach to human behavior. Sir Francis Galton, James McKeen Cattell, and Alfred Binet were pioneers in the test and measurement field.

The originators of intelligence tests assumed that intelligence is an innate ability that individuals possess more or less of in relation to each other. The growth of intelligence was described in terms of age norms. The underlying assumption was that as the body grows, so does the mind. Just as the "average" 4 year old is so tall and weighs so many pounds, the 4 year old of "average" intelligence knows and can do so many things. Moreover, just as some are tall and some are short due (in large part) to their genetic makeup, it was assumed that some are more or less intelligent due to their genetic makeup. Although Darwin's influence is apparent, this approach to the study of intelligence is not based on a formal theory of intelligence per se. Also, no attempt was made to specify the processes by which intelligence grows and changes. Tyler (1976) described the era in which intelligence testing began as follows:

> It was a time in which differences in the general quality and worth of individuals, classes and races were assumed without question...The early mental testers did not originate the idea of hereditary differences in intelligence, they simply accepted it. Assuming that innate differences existed, they saw their job as one of measuring the differences as accurately as possible. It is quite possible that the testing movement accelerated the breakdown of classes by identifying able individuals from the "lower" strata who might otherwise have gone unnoticed. (P.15)

(2) Behavioristic approach. Psychologists who adopted a behavioristic approach to learning and intelligence limited their analyses to overt and measurable behavior. The ability-performance issue was avoided by focusing on performance only. This was seen as a way to remain scientific. Non-observable entities such as thinking were not studied.

According to this school of thought, intelligence is the sum of prior learning experiences. We cannot be sure whether or not people differ in intellectual ability, but we do know that they differ in their learning histories. Thus, if two people differ in intelligent behavior, it was thought to be due to differences in prior experiences rather than innate abilities. Behavioristic psychologists were content to study animals such as rats or pigeons, since they wanted to describe the laws of learning and behavior, which were assumed to be applicable across species. In the United States,

this approach was popular from the 1920s to the 1960s. Some leaders in the field were James B. Watson, Clark C. Hull, and B.F. Skinner.

(3) Developmental approach. Although it did not become popular until the 1960s, the developmental approach to the study of intelligence also began at the turn of the century. In addition to describing what children at various ages knew and could do, developmental psychologists attempted to explain changes in intelligence over the course of life. To accomplish this objective, psychologists not only studied children of different ages, but they also compared human performance to that of animals. Thus, intelligence was viewed from an evolutionary perspective which considered phylogenetic as well as ontogenetic changes. Developmental psychologists were willing to posit and study nonobservable entities such as thinking. They did not view intelligence as innately given, nor did they view it as the additive sum of past experiences. Instead, developmental psychologists maintained that there are qualitative differences in thinking at different periods or stages of life. For example, an 8-year-old child not only knows more about the world than a 3-year-old child does, but he or she also thinks and understands the world differently than the younger child does. Leaders in this field during the early part of this century were Jean Piaget, Lev. S. Vygotsky, and Heinz Werner.

Each of these three psychological approaches to the study of intelligence exists today. During the past 20 years other approaches have also become popular. Computer-based artifical intelligence and information processing are the most notable of the newer approaches. In the field of psychology, then, ideas and theories about intelligence as well as methods of studying intelligence are quite varied and have changed over the past 80 years.

The various approaches to the study of intelligence have each had an impact on educational practice. For example, Skinnerian learning theory lead to the use of programmed instruction in schools. Similarly, several school programs are said to be based on Piaget's developmental theory. However, of all the approaches to intelligence, the test and measurement view has had the largest impact on education, as its methods have become an integral part of the schooling process itself. In the next section we will consider why this came about.

TESTING AS A SCHOOLING PRACTICE

In order for intelligence tests to be understood, it is helpful to appreciate why they were constructed in the first place. Perhaps the most important factor leading to intelligence testing was the perceived need for such tests in school. Intelligence tests are rooted in the schooling process (Resnick 1976). The first measures (or tests) of intelligence were designed to determine which children would have difficulty learning in a typical classroom environment.

Some 75 years later, this is still one of the primary functions of intelligence tests--to predict academic achievement. As it turns out, the tests do this rather well. If you wanted to estimate or predict how a child entering first grade will perform in school and you could only use one piece of information about that child, you would be more accurate if you used his or her score on an intelligence test than if you looked at the child's age, sex, ethnic background, socioeconomic status, and so on. Since schools still feel the need to make predictions about academic achievement and since intelligence tests do a good job, they are widely used. In fact, although the measures of general intelligence used today have been revised and upgraded, they are

not that different from the first version used in the early 1900s (Horn 1979).

Schools existed long before intelligence tests. You might ask then, when and why did the need to use tests to predict school success arise? As with many other school practices, intelligence testing began in response to a demand placed on schools by the public. This point is often overlooked. To appreciate why intelligence testing became a schooling practice, we should review changes in our society's expectations of schools. For the sake of illustration, three points are presented below.

(1) The first point pertains to the establishment of schools and a delineation of their functions--in other words, the acceptance of the idea of formal education itself. It benefits society to create institutions in which instruction and evaluation of pupils occur. Education is conceived of as a process of effecting progressive change in individuals. The major purpose of education is to transmit the knowledge and values of society to its youth in order to prepare and screen individuals for a role in society. Schools are asked to instruct students and to evaluate their progress in benefiting from that instruction. Results of the evaluation process are used in conjunction with other information to place persons in occupations. As Bronowski (1973) pointed out, one of the great achievements in cultural evolution was the idea that a boy did not have to assume the trade or occupation of his father. Schools were one important vehicle that allowed this flexibility and opportunity for advancement to take place.

(2) A second point is that, in time, the public decided that all children should be entitled to an education. In the middle of the last century, systems of universal compulsory education were mandated. Such laws were

certainly consistent with our country's heritage of democracy and the idea of equal opportunity for all citizens. By 1918 all states had a requirement that their resident children must attend school for a specified (minimum) period of time. As Tyler (1976) has noted:

> The historical fact is that the demand for intelligence tests arose everywhere in the period after school attendance was made compulsory. Obviously what the legislation did was to assemble in one place, probably for the first time in human history, almost the full range of human intellects, and to make it necessary for educators to struggle with this diversity. Society has not yet met this challenge. People still blame the schools or tests for the magnitude of the variation that compulsory attendance reveals. The first mental-test developers and their successors were at least attempting to cope with the problem (P. 14).

Intelligence tests were the schools' response to the challenge of wide individual diversity in the classroom. It became apparent that many pupils could not succeed in a "regular" classroom. Rather than change the classroom which worked for the majority of students, schools decided to screen out of the regular program those children who could not learn like the others. In this way schools could reduce the number of failures and increase their efficiency. For how could schools be held accountable for the failures of children who were mentally defective? Some children would have to have their needs met in special programs rather than in the regular school programs.

Thus, by sorting students into different programs, schools attempted to maximize their successes and meet the needs of the learner who was atypical.

But which children were "special"? Until intelligence tests arrived, such classifications were based on teachers' or medical doctors' judgements. This was felt to be inadequate and unfair to many children since these determinations were affected by factors other than intelligence per se (such as social and family background, appearance, interpersonal style, and so on). At the turn of the century, Alfred Binet set out to devise a more objective way to assess whether a child could be expected to benefit from the average Parisian classroom. By 1920, Binet's test (in modified form) was being used widely on this side of the Atlantic Ocean as well as in Europe.

(3) As a final point, consider that society has recently placed a new demand or expectation on its schools. In the past, schools were expected to offer instruction to their pupils using curricula developed by the educators and approved by parents. The perception has grown that many students have not mastered their school's curriculum even though they are promoted to the next grade or awarded a diploma certifying graduation. Without reviewing the evidence as to whether this perception is accurate, suffice it to say that it has resulted in a call for new proof that pupils are not only exposed to instruction, but that they learned something as a consequence of going to school. In other words, schools are being held accountable for the outcome of the instructional process. Schools are being told that they must demonstrate that their "clientele" have mastered at least parts of the standard curriculum. To meet this public demand, some states have implemented minimal competency tests as part of the exit requirements for promotion and/or graduation.

As was the case in Binet's time, the teacher's judgment is being supplemented by a test. It is not enough that the teacher gives a child a passing

grade. The child must also pass a statewide test. As was also true in the time of Binet, this movement toward testing is viewed by most concerned parties as progress. An objective measure applied to all individuals will determine whether they have been properly educated. Thus, rather than move away from testing as a schooling practice, our society has increased the use of tests in the process of schooling. This point will be returned to at the end of the paper.

Summary. This section described testing as a schooling practice. Intelligence tests were developed around the turn of the century. The major purpose of the tests was to identify individual differences in intelligence which were presumed to cause differences in school achievement. The tests were more scientific and objective than the previous methods used to identify differences among children's intelligence. A need to identify individual differences in preparedness to learn arose when education became compulsory. In recent years, communities have asked that schools be held accountable for the outcome of the instructional process. To meet this need, many states have implemented minimal competency tests as a requirement for promotion or graduation. Thus, the schooling practice of testing children may be viewed as a response to the demands placed on schools by the public. For this reason, it is important that the public be clear about the strengths and limitations of the tests and the use of tests. The next section provides further details on the origins of intelligence tests.

BINET'S SEMINAL WORK

We will review Binet's work in order to further clarify the nature of intelligence tests and their relation to schools. Binet set out to devise a

mental test that could be used to identify those children who would not be able to benefit from a public education in Paris around 1900. Binet constructed his test by presenting different problems to children who were judged by teachers and whose school performance showed them to be "good" or "bad" students in terms of their academic achievement. Children were tested individually on several types of problems. Those problems on which the better students outperformed the poorer students were retained for subsequent development of the test.

Notice the logic of this approach. The goal is to create a test that allows for estimation of who will not be able to learn in school. You begin by taking children whose school performance is already known. You find problems on which the students doing well in school outperform those doing poorly in school. You then set out to further refine those items so that they can be used with other children to predict how well they will do in school. If a child performs on the test in a manner similar to one of the better students, that child is thought to have the intelligence necessary to learn in a regular classroom. One essential feature of the Binet scale is evident here. School performance was used as a criterion for intelligence. Those who cannot learn in school are considered to be less intelligent than those who can learn in school. This assumption inherent in intelligence testing has been called scholastic validity (Furth 1973).

It was not the case that any problem Binet gave served to differentiate among students. In particular, Binet found that the sensory and special ability items used by Cattell in 1890 with college students did not distinguish the good from the poor students that he tested in Paris. Binet then tried

problems that required higher-level mental processes such as attention, comprehension, and memory. These items "worked" and were therefore retained for the test.

Once the types of items were identified, Binet and his co-worker T. Simon further refined the test by administering it to scores of children aged 3 to 15. Based on the children's performance, the test items were rank ordered and grouped by age levels. A test was considered appropriate for a given age, say 4 year olds, if 65 to 75% of the 4 year olds could pass the test. The original sample of children is called the norming group. This group sets the standards for future children who will take the test. A child's performance is either below average, average, or above average in relation to the original norming group. Thus, two more features of Binet's work are evident. First, the items on his measure of intelligence were derived through a trial-and-error process. The items were not based on an existing theory of intelligence. Instead, the items that "worked"-- i.e., those that distinguished among children of different ages--were retained. Second, Binet's measure of intelligence was norm-referenced. A score on the test reflected a measure of performance which was relative to that of the original pool of subjects.

The final version of the scale developed by Binet and Simon contained 54 tests. These are listed in the Appendix according to age levels. Although at each age level the tests measure different things (math ability, memory, and so on), a child's performance on the test was used to derive a general or global level of intelligence called mental age (mental level was the term used by Binet). To simplify somewhat, if a child passes all the items up to 8 years and goes no further, that child's mental age is 8. To say that a

child's mental age is 8 is to say that he or she did as well as the original group of 8 year olds upon whom the tests were normed. Another feature of Binet's scale, then, is that it yielded a global index of intelligence called mental level or mental age.

The Binet-Simon scale was soon adapted for use in the United States by Lewis Terman of Stanford University. In 1916 Terman published the Stanford-Binet Intelligence Scale which was normed with American children. This test was revised and renormed in 1937 and again in 1960. Binet's original test has been both a model and a criterion for subsequent intelligence test construction. It has been a model because many current tests of intelligence are similar to Binet's in that they (1) are individually administered, (2) are norm-referenced, (3) are standardized, (4) contain subtests, and (5) yield a global index of intellectual level. Binet's test has been a criterion because subsequent tests have had to show that they yielded results similar to that of Binet's in order to demonstrate their validity. Since most intelligence tests are validated by showing that their results match those obtained with prior tests, it is important to remember Binet's original method of test construction. The next section summarizes properties and types of modern intelligence tests.

PROPERTIES AND TYPES OF INTELLIGENCE TESTS

The intelligence tests reviewed here are similar in that they all derive from the test and measurement approach (called the psychometric approach) to intelligence. The developmental and information processing approaches to intelligence have come up with methods to measure intelligence, thinking, and problem solving, but these methods are not covered in this paper.

Let us begin with two properties that all intelligence tests have in common. They are standardized and norm-referenced.

(1) Standardized tests. To say that a test is standardized is to say that the conditions under which the test is administered are clearly prescribed so that they are the same for everyone who takes the test. For example, a specified period of time might be allowed for a given part of the test. (Remember? "Time is up. Put your pencils down now!")

The idea behind standardization is that those who take the test are doing so under the same conditions as those who were used in the original test formulation. For if the conditions of test administration vary in significant ways, the test will not be measuring the same thing across persons. Of course, standardization is an ideal that is difficult to obtain in practice. For example, on an individually administered test (see below), the examiner and examinee are not always matched for sex and ethnic background. If these kinds of differences cause some examinees to feel uncomfortable when taking the test, their performance may not be as true a measure of ability as the performance of someone who is at ease when taking the test. Skilled examiners can recognize and attempt to minimize such extraneous performance factors.

(2) Norm-referenced tests. Intelligence tests are also norm-referenced. This means that the test score provides a relative measure of performance. It tells you how you did relative to those who were used in the original test development. "You are above average for you age" is an example of a norm-referenced statement. It is extremely important, then, to know who the norming group is in order to interpret this kind of statement. For example, suppose you were told that Jim is the best bowler in the group. Does this

tell you how good a bowler Jim is in any absolute sense? The answer is no. Jim might be part of a fifth grade team and his average might be 97. Conversely, his average might be 197, and he may be on the professional tour.

The reputable intelligence tests of today were normed on thousands of males and females of different ages who were "normal," i.e., not blind, deaf, or otherwise evidencing a physical disability. Persons from different ethnic, racial, and social class backgrounds were used to norm the tests. Selection of different backgrounds is done such that it matches or is proportional to the composition of the population as a whole. People who form a norming group are a "representative sample of people who are similar to the people for whom the test is intended" (Smith 1979).

Intelligence tests, then, do not provide an absolute measure of ability. The test results are meaningful only to the extent that the person tested is similar to the norming group. For example, if you do not understand the language of the test, you will probably perform at a below average level because the test was normed with and intended for normal speaker-hearers. Similarly, if you memorize the correct answers to an intelligence test, you will score way above average since you will outperform those who originally took the test without prior knowledge of the answers. In sum, your score on an intelligence test reflects how well you did relative to the norming group. The distribution of scores on intelligence tests is reviewed in the next section of the paper.

Next we consider some properties that tests do not have in common and which result in different types or kinds of intelligence tests.

(3) Individually administered tests. As the name implies, these kinds of tests are taken by one person at a time. The tests are administered by a trained examiner, i.e., someone who has been taught how to give the tests. Examples of individually administered intelligence tests are (a) the Stanford-Binet Intelligence Scale, Form L-M; (b) the Wechsler Preschool and Primary Scale of Intelligence (WPPSI); (c) the McCarthy Scales of Children's Abilities; (d) the Wechsler Intelligence Scale for Children--Revised (WISC-R); and (e) the Wechsler Adult Intelligence Scale (WAIS).

These individually administered intelligence tests consist of subtests or subscales such as a section on vocabulary or arithmetic. The items require the examinee to answer verbal questions, solve problems, remember presented information, and so on. Many of the tests require the examiner to record the time it takes for the examinee to complete the test sections. If a test is timed, then a person's answer coupled with the time it took to derive the answer are used to obtain a score. Over the past 10 years the WISC-R has become the most frequently used measure to assess school-age children's intelligence. The McCarthy Scales are becoming widely used to assess preschoolers' intelligence. Individually administered intelligence tests are often used as clinical tools. The reader is referred to Klein (1977) for a discussion of this use of intelligence tests.

(4) Group-administered tests. Since individual tests require a trained examiner and assess one person at a time, they are expensive and time-consuming from a practical standpoint. If you wanted to test hundreds or even thousands of people, administration of the Stanford-Binet or WISC-R would take a great deal of time and be very expensive. To meet the needs of this practical problem, group-administered tests were constructed.

The first successful use of group intelligence tests in the United States occurred during World War I. The Army Alpha (for English speakers) and Beta (for non-English speakers) tests were used to place thousands of new recruits in military programs such as officer training. Today, group tests are frequently used by schools as entrance requirements and for program placement. An example is the Scholastic Aptitude Test (SAT) which is required by many colleges as part of the application packet.

Unlike individual tests, group intelligence tests are paper-and-pencil tests. They typically consist of subtests or parts, each of which is timed. There must be a person to administer and proctor the test for the group, but this person does not need the kind of training that an individual intelligence test examiner needs.

(5) Global versus specialized measures. Both individual and group tests typically consist of subtests. For example, the WISC-R consists of a verbal section and a performance section. The subtests for the verbal section are called information, vocabulary, similarities, arithmetic, and verbal comprehension. The subtests of the performance section are picture completion, picture arrangement, block design, object assembly, and coding. A person is given a score on the separate subtests. The scores are also combined to yield a global index of intelligence.

There are tests that yield scores for separate and specialized abilities only. An example of a specialized ability is spatial ability. An example of such a test is the Differential Aptitude Test (DAT). The different measures of intelligence (global versus specialized) reflect different theoretical

conceptions of human intelligence. Since tests of intelligence first appeared, psychologists have debated whether or not human intelligence is best described as one general ability or as separate abilities. Today, there are several competing models of human psychometric intelligence. Some give primacy to a general ability, whereas others emphasize separate, relatively unrelated abilities. Thus, the fact that there are general and specialized measures of intelligence stems in large part from differences among psychological models of human intelligence.

Today, there are individual and group tests that yield both global and specific measures of intelligence. All four kinds of tests are used in the process of schooling and in clinical diagnoses (to determine why a child is having difficulty learning in school, for example). The reader is referred to Klein (1977) for a more complete review of types of intelligence tests including those designed for special children. A more technical reference on all types of tests is Oscar K. Buros' The Mental Measurements Yearbook. This source contains critical reviews on most existing tests. If you want to find out what a test is, what is purports to measure, its norming group, its evidence for validity, and so on, Buros is an excellent reference.

(6) Measures of ability versus achievement. Another dimension on which tests differ is the extent to which they measure intellectual ability or achievement. It is clear that intelligence tests were thought of and designed to assess an ability (aptitude or trait). An examination of the content of intelligence tests shows that many items are not explicitly taught by the culture. To the extent that items are not taught, they may be considered to be measures of ability and not achievement. Moreover, since performance on

intelligence tests is positively correlated with performance in other situations said to require intelligence, there is evidence that the tests are measuring an ability that is stable over time and across situations. In particular, the tests do a good job in predicting academic and occupational achievement.

An argument can be made, however, that intelligence tests are measuring achievement rather than aptitude. Some of the test items are explicitly taught in school or at home. Gage and Berliner (1979) state that intelligence test questions differ from achievement tests only in the degree to which they measure things that are taught in school. As for prediction, it turns out that achievement tests do a good job of predicting academic performance also. For example, if you wanted to estimate how well students will do in tenth grade geometry, their nineth grade algebra performance will be an excellent predictor. Thus, the fact that intelligence test performance is related to performance in other situations is not a guarantee that the tests measure ability and achievement. The distinction between aptitude and achievement is easy to make in theory but difficult to realize in practice. It is a source of controversy in the field of educational measurement that is far from being resolved (Smith 1979).

You might ask why this distinction is important in terms of intelligence testing. Even if the tests have "achievement items," won't those persons with greater ability master the items faster than those with lesser ability? In fact, this is one of the underlying assumptions of intelligence tests. Achievement is viewed in large part as a function of aptitude. To illustrate this point, let's take an example from the field of athletics. If we observe that two people receive the same set of beginning tennis lessons and that at

the end of the lessons, one is a far better tennis player than the other, we will probably conclude that one has more "natural athletic ability." Of course, we recognize that differences in tennis achievement could be due to nonability factors such as motivation to learn the game, effort, prior athletic experiences, and so on. We recognize that aptitude is one of several factors underlying achievement.

Now, the problem with making inferences about differences in ability based on intelligence test performance is that it is clear that everyone does not receive "the same set of lessons." The norming procedure accounts for these differences in achievement factors but does not eliminate them entirely. In addition, the norming group is more representative for some individuals than it is for others. Environmental factors that might affect intelligence test performance are nutritional history, stability of home environment, amount of intellectual stimulation received, and so on. In short, for many individuals an intelligence test may in fact be a valid measure of ability (which determines achievement). For others, we may be erroneously making ability inferences when the performance observed is more a function of environmental factors. For example, a child may score poorly on a vocabulary section on an intelligence test because his or her environment has not presented him or her with the lexical items that appear on the test.

In terms of schooling practices, the inability to distinguish between the ability and achievement components of intelligence test performance is important to recognize. For if a child has a "below average" score, we cannot be certain whether this reflects a lack of opportunity to learn what is required on the test, or whether it instead indicates a slow rate of picking up information

from the environment due to learning difficulties (Resnick 1979). The former problem is one of achievement; the latter is one of ability. Since schools have typically assigned students to different instructional programs based on their measured ability, it is important that an accurate diagnosis of ability be made. A student who is incorrectly placed in a low-ability program may never realize his or her full potential. We will return to this point later on in the paper.

INTELLIGENCE TEST SCORES: WHAT IS IQ?

Recall that an intelligence test yields a norm-referenced measure of intelligence. The score provides an indication of how a person did relative to the group used to norm the test.

Technical details on how intelligence tests are scored will not be presented here. To simplify, test responses are scored as fully correct, partly correct, or incorrect. A certain number of points is awarded for each partly or fully correct answer. For many tests, the time taken to answer a question is recorded along with the answer itself. The time is used in conjunction with the answer to score a response. Scores are summed to yield a raw total score for the test, or in most instances, for the various subtests that comprise the test. In turn, these summed scores are converted into a standardized score that is used to compare individuals who took the test. One such standardized score is intelligence quotient, or IQ. IQ, then, is a score on an intelligence test. It tells you how an individual performed on the test in relation to the standards set by the norming group.

Although the methods for computing IQ have changed over the past 20 years, it is interesting to note the history of this scoring technique.

Scores on the Binet-Simon scale were expressed in terms of mental age. Terman (1963) realized that this score was not very informative. For three children could each be found to have a mental age of 8 but they might be 6, 8, and 10 years old in terms of chronological age. Since the test items were scaled by age, the first child would be above average in comparison to other 6 year olds, the second child would be an average 8 year old, and the third child would be below average for 10 year olds.

Terman decided that it would be more informative to express the results of an intelligence scale in terms of a relation between mental age (MA) and chronological age (CA). That is MA divided by CA. For ease of expression, the ratio is multiplied by 100. Such a score was called intelligence quotient, or IQ. The formula for IQ was as follows:

$$IQ = \frac{\text{mental age in months}}{\text{chronological age in months}} \times 100.$$

In the above examples, the IQ of the three children would be 96/72 x 100 =123; 96/96 x 100 = 100; and 96/120 x 100 = 80. (Note that the children were assumed to be exactly 6, 8, and 10 years old respectively, and exactly 8 years in terms of mental age.) Thus, an IQ of 100 is average, as mental age matches chronological age. Scores above 100 reflect above average performance, scores below 100 reflect below average performance.

The above formula for computation of IQ is no longer widely used. Tests such as the Wechsler Scales come with scoring charts that allow the examiner to convert raw scores into standard scores for each subtest, and in turn, for the whole test. The logic underlying the chart norms is identical to that which underlied Terman's formula. The score tells you how an individual performed relative to same-aged "peers" who were used to construct the test.

INDIVIDUAL AND GROUP DIFFERENCES IN IQ

Having shown what IQ is, let us now consider how people actually score on intelligence tests. Three points are considered in this section: (1) how a large number of people would perform on tests such as the Stanford-Binet or WISC-R; (2) the nature of individual differences on intelligence tests; and (3) the nature of group differences on intelligence tests.

(1) Performance of groups on IQ tests. We have seen that an IQ of 100 is average. The distribution of IQ scores about this average is normal. (The normal distribution is a technical term in statistics. Many human characteristics have been found and/or are thought to be normally distributed.) For example, suppose the WISC-R were administered to thousands of children aged 6 to 16. The following results would be obtained. Approximately 50% of the people tested would score between 90 and 110, and approximately 16% would score between 80-90 and 110-120. In other words, 82% of the population would have an IQ score of 80 to 120. Only 9% would score below 80 or above 120. Administrations of IQ tests to groups of people have confirmed these expectations. The labels attached to ranges of IQ scores on the Stanford-Binet and Wechsler test are listed below.

IQ Range	% of Population	Stanford-Binet	Wechsler
140 & above	0.4	Very Superior	Very Superior
130-139	2.0	Superior	Very Superior
120-129	7.0	Superior	Superior
110-119	16.0	High Average	Bright Normal
90-109	50.0	Average	Average
80-89	16.0	Low Average	Dull Normal
70-79	7.0	--- Borderline Defective ---	
60-69	2.0	--- Mentally Defective ---	
59 & below	0.4	--- Mentally Defective ---	

(2) Individual differences on IQ tests. How stable over time is a person's IQ score? Studies of the same people over time have shown that an IQ score does not become stable until a person is about age 10. (Even at age 10, IQ is not fixed, although it is likely to be very close to the person's IQ as an adult.) Below age 10, the relation between childhood and adult IQ has been found to vary in both level and direction of differences. For example, changes in IQ scores of 15 points between ages 6 and 10 occur about 20% of the time. The scores may increase or decrease. Note that such a change would cause a "label" shift of "borderline defective" to "average," of "average" to "superior," and so on, and vice versa. Changes in intellectual "ability" are even more dramatic when performance on tests designed for infants are compared with later performance on tests such as the WISC-R. In general, there is little, if any, relation between performance on infant and childhood IQ tests. Thus, research has shown that a person's IQ score is not fixed and unchangeable.

Leaving the question of stability aside, let us now consider why it is that one person's IQ is 120 while another's is 85. What are the causes of individual differences in IQ scores? This is a complex issue which psychologists have been debating since the field first began. The broader question is called the "nature-nurture controversy." Are people different because of their genetic backgrounds, their environmental experiences, or a combination of the two?

To address the issue of the causes of differences in IQ scores, researchers have conducted studies of people whose genetic and environmental backgrounds vary in systematic ways. For example, IQ scores of identical twins (who have the same genetic code) reared together have been compared with IQ scores of

identical twins reared apart. Identical twins reared at home versus non-twin siblings reared at home is another example of the kinds of comparisons that have been made. In general, the results of this research are interpreted as demonstrating that individual differences in IQ within one particular social group are more due to genetic than to environmental factors (Gage and Berliner 1979). For this research has shown that as similarity of genetic background increases, so does similarity in intellectual performance. Moreover, it is recognized that many cytogenetic diseases are associated with below average intelligence test scores. Thus, there is evidence that one's genetic makeup plays a role in shaping one's intelligence test score.

(3) Group differences on IQ tests. The nature-nurture issue crops up again and becomes more controversial when group differences in IQ scores are considered. The fact of group differences is not disputed. For example, numerous studies have found that the average score for Caucasians is about 15 points higher than the average score for blacks on IQ tests. Similarly, on a specialized ability such as spatial ability, males have been found to outperform females in terms of average level of performance. What is controversial about such group differences in intellectual performance is the explanation of them. There is not agreement among experts in the field as to whether group differences are mainly due to genetic or environmental factors.

Although individual differences within one group are mainly due to genetic factors, this does not imply that differences between groups are also due to genetics. The view that one group is genetically inferior to another in terms of intellectual capacity is not widely accepted today. Those who adhere to an environmental explanation of group differences point to differences

in rearing patterns, opportunities, and so on to account for the observed differences.

A point on group differences is important to apprehend. The difference, say between whites and blacks, on IQ tests is an average score difference, not a difference in distribution of scores. For each group, the distribution of scores is normal. The "average white score" is about 15 points higher than the "average black score." But for both white and black groups, IQ scores range from below 40 to above 170. In other words, there is much overlap and similarity between the distribution of white and black group IQ scores. In fact, there is more overlap than there are differences. The difference between the lowest-scoring white and the highest-scoring white is much greater than the difference between the average white and the average black. Many feel that these differences within groups are more important to account for than are the differences between groups.

Disagreements concerning the causes of individual and group differences on IQ tests have led to two challenges or questions of the tests. First, do the tests really measure intelligence? Second, are IQ tests biased in favor of certain groups?

DO IQ TESTS REALLY MEASURE INTELLIGENCE?

Since definitions of intelligence reflect cultural values to some degree, it is difficult, if not impossible, to create a culture-free measure of intelligence. If we cannot define intelligence universally, how can we hope to devise a test that is universally applicable? Note that the idea of creating an intelligence <u>test</u>, regardless of its content, reflects a cultural bias in and of itself.

Given this caveat, we may ask whether intelligence tests constitute a valid measure of intelligence as defined by our culture. (Few today would accept E.G. Boring's view that intelligence is whatever intelligence tests measure.) There are rather diverse differences of opinion on this topic. Some, such as Bereiter (1976-1977), assert rather strongly that IQ tests are valid measures of intelligence. One of Bereiter's arguments is quoted below.

> Define intelligence any way you please. Now, given two randomly selected people, which one is likely to prove more intelligent (according to your criteria, whatever they are), the one with the higher IQ or the one with the lower IQ? By any but the most perverse criteria of intelligence, the person with the higher IQ tends also to be higher on your criteria of intelligence. The reason for this robustness of prediction is that IQ has been correlated with more things than any other psychological variable by far, and with hardly any exceptions it is found that higher IQ scores are associated with more favorable standings on other variables (P. 38).

In other words, the evidence for IQ as a measure of intelligence is that it is a good (if not the best) predictor of performance in numerous other situations that are said to require intelligence. Such predictions are made in terms of groups. Naturally, there will be exceptions in individual cases.

One important setting to which IQ scores have been shown to be related is school performance. While not disputing the findings of a positive relation between IQ and academic achievement, some have questioned whether success in school does, in fact, require intelligence. In other words, is high academic

achievement a valid criterion for high intelligence? Does poor school performance reflect a lack of intelligence? Although there are no pat answers to these questions, it is clear that a distinction can be made between academic and everyday thinking (Brown and French 1979; Neisser 1976). Academic problems tend to be abstract and to require speed, and they usually have a definite answer and method of solution using given information. In everyday problem solving, in contrast, speed is often irrelevant; there is usually a strong emotional involvement in a specific outcome; and often the necessary information is not at hand (Brown and French 1979).

This distinction between academic and everyday thinking does not deny that intelligence tests measure intelligence. It does highlight the fact that the tests have school performance as their major source of validation. In general, schools require and focus on verbal abilities, such as reading and speech, and on the ability to solve abstract problems, such as those found on classroom tests. These abilities are stressed in school because they are valued in our culture. Since IQ tests assess the degree to which one has and/or will master these abilities, they are considered measures of intelligence in our culture.

ARE IQ TESTS BIASED?

The fact of individual and group differences in average IQ scores, especially between black and white children, has led to the question of bias in the tests. This is another complex issue which cannot be fully treated in this paper. There are two separate aspects to this question. One leads to the conclusion that the tests are not biased, and the other points to a clear bias.

On the one hand, measurement specialists have analyzed the technical qualities of the tests across different groups. One important finding is that IQ tests predict academic achievement for groups of black children as well as they do for any other group. Using school performance as a criterion, IQ scores work well for both black and white children. Thus, blacks with higher IQs tend to outperform blacks with lower IQs in school. In addition, technical analysis of item performance on IQ tests shows little, if any, difference between groups of whites and blacks. To simplify, both white and black children have been shown to find the same items to be easy relative to other test items that are more difficult. For these reasons, it is argued that IQ tests are not a biased measure of intelligence. They work well for most groups in terms of prediction, and they distinguish among children in similar ways in terms of item difficulty.

On the other hand, it has been pointed out that IQ tests measure skills and facts that are more accessible to certain social classes than they are to other social classes. Children from middle and upper classes are more likely to have an opportunity to learn what intelligence tests measure. A low test score does not distinguish between whether a child lacked opportunities to learn or indeed is slow in learning. This distinction is critical, especially when the performance of children from educationally disadvantaged populations is being considered (Resnick 1979).

The argument for bias in IQ tests shifts, then, to a consideration of how the tests are used. (This is what the courts are now trying to decide. What is the appropriate use of IQ tests?) Specifically, the controversy centers around how the lower scores of certain groups are interpreted and translated

into school practices. Many teachers and school administrators think of IQ scores (and the intelligence they reflect) as absolute constants. They feel there is little that can be done to improve a child's intelligence since this is an innate, fixed ability. Based on a low test score, a grade school child may be placed in a "slow" or remedial class and treated as if he or she could not learn or could learn only very slowly. Such treatment may serve to guarantee that the child will not learn and will thereby remain below average on the abilities required in school and on intelligence tests (Gage and Berliner 1979). Note how unfair this is if the original low IQ score was due to a lack of opportunity to learn what the majority culture considers to reflect intelligence rather than to a lack of ability per se. Moreover, recall that children's IQ scores are known to vary, to not be fixed and unchangeable in many cases. The scores fluctuate considerably from early to middle childhood. A child who is "mildly defective" at age 6 might be "average" at age 7.

Recognition of the self-fulfilling prophecy outcome of IQ-based classifications in schools has led to the call for banning the use of intelligence tests for school program placement. The perceived misuse of IQ tests has also led to a search for alternatives to intelligence testing. This issue is considered in the final section of the paper.

ALTERNATIVES TO IQ TESTING

Are there alternative ways to measure intelligence other than by using IQ tests? As we have seen, the answer is a clear and definitive yes. Over the past 70 years psychologists have identified numerous intellectual processes that are not directly assessed by IQ tests. In addition, alternative views of

intelligence and its development, not based on age norms, have been proposed and supported by psychological research. These alternative views of intelligence have led to alternative kinds of tests of intelligence. As far as children's intelligence is concerned, Jean Piaget's theory is a prime example of a different theory of intelligence which has different measurement techniques (De Lisi 1979). As Horn (1979) has stated, the IQ tests of today are very similar to Binet's first tests. The technology of assessing intelligence has not kept pace with what we now know intelligence to be. Horn feels that the failure to substitute more sophisticated measures of intelligence is a matter of inertia. That is, it is difficult to change the technology of measuring intelligence.

It seems that there is another, equally important reason for the persistence of intelligence tests. The same tests are still used because the same school practices are still in vogue. In other words, the search for alternatives to IQ testing must go beyond the tests themselves to a consideration of school practices and policies. Rather than search for new ways to measure intelligence, we need new ways to meet the needs of individuals in school (Glaser 1976; 1977).

Consider the functions that IQ and similar kinds of tests serve in the schools today. According to Resnick (1979), there are three major functions.

(1) Tests are used to facilitate and manage the instructional process by sorting pupils into different programs.

(2) Aptitude tests are currently serving an accountability function. The public is monitoring and putting pressure on schools to increase their productivity. Aptitude test results (such as SAT scores) are used to indicate a school's degree of success.

(3) Finally, standardized tests serve to legitimate the schooling process. They help to make the schools more scientific by providing an independent and objective measure of aptitude and achievement.

As Resnick (1979) has also pointed out, the increasing demands for school accountability may ensure that IQ tests remain on the scene for some time. Schools may point to IQ test scores as a way to place limits on public expectations and demands. Similarly, there is a movement away from using teacher-based judgments as the sole criterion for promotion and graduation. Statewide minimal competency or basic skills tests have recently appeared in several states. These tests are supposed to monitor the outcome of the schooling process, in an objective and scientific manner, to preserve standards. Thus, in terms of points 2 and 3 above, rather than moving away from using IQ tests in school, we seem to be increasing the demands for such a practice. Even those who call for the abolition of IQ testing in school are probably not opposed to the notions of schools' being accountable to the public and being objective and scientific in making decisions about student performance and student placement.

The search for alternatives to IQ testing in schools, then, must center around point 1 above, that is, using tests to manage the instructional process by sorting pupils into different programs. Rather than being on the decline, this practice has also had a resurgence. Many school districts are using IQ tests to help determine whether a child is "intellectually gifted" and should, therefore, be placed in a special program. Note how consistent this is with the tradition started by Binet. Schools sort pupils into different programs (remedial, average, or advanced) to better manage instruction and thereby

maximize their resources. In other words, intelligence tests are a way for schools to meet the needs of individual pupils. The tests identify differences among children and are used as part of an evaluation process to determine which program will best suit a particular child's needs.

Why has this practice persisted and even been on the increase of late? The answer is rather straightforward. For the vast majority of children, the system works rather well. Moreover, the present model of schooling is consistent with our society's view of what comprises good education. Most parents want schools to evaluate and compare children's ability and performance, and they want this process to begin early in grade school. As a result, first graders are compared on reading tests in the same way that high school seniors are compared on the SATs. The first graders who are better at reading are put into more advanced second grade reading programs than their lower reading-level peers. Similarly, those with higher SAT scores are admitted to better colleges than those with lower SAT scores (all other things being equal). Since our society is competitive, the schools are charged with preparing children for this fact. The idea of eliminating competition and evaluation in school would probably not strike a responsive chord in our society today.

Banning IQ tests will not ban individual differences in children's ability and readiness to learn in our schools. In fact, banning IQ tests and maintaining our present model of schooling might bring us back to the era before Binet. In short, if we want to find alternatives to IQ testing, we need to identify alternative ways for schools to meet the needs of individual pupils.

Over the past 70 years, the search for ways to improve schooling vis-a-vis the needs of individuals has persisted. A comprehensive review of the many alternatives that have been proposed is beyond the scope of this paper. The following are only examples of the many alternatives that have been offered.

Glaser (1976; 1977) has described the present model of schooling as a selective, fixed-track approach. Pupils are sorted into programs based on ability (as measured by tests) and prior achievements. Thus, schools admit to their programs those students who are known to have the abilities required to benefit from their instructional methods and thereby eventually meet their exit requirements. In other words, the system is designed to maintain long-standing instructional techniques and evaluation systems. Rather than accommodate different styles of learning, the schools insist that all students must fit the same academic mold.

Glaser points out that the present method of schooling is based on ideas and principles from educational psychology which are over 75 years old. We now know that ability is not fixed and unchangeable during a person's lifetime. Thus, to track a child into a program based on tests of ability, especially when the child is under 7 years old, flies in the face of what we know about the stability of that ability. Moreover, we now know that different styles of learning exist which are not adequately considered in current instructional techniques. The reader is referred to Glaser (1977) for a review of alternative systems of instruction and evaluation that are designed to meet the needs of individual pupils in schools.

Another approach to meeting the needs of individual learners is called mastery learning. The basic idea behind mastery learning is that children can attain comparable levels of achievement if given a sufficient amount of time

to learn. In the traditional schooling approach, a set amount of time is given to all children to master a lesson. Naturally, those with greater ability are at an advantage relative to their less able counterparts. This approach guarantees that those with greater ability will tend, on the average, to achieve at higher levels. A mastery learning approach calls for manipulation of time to learn in accordance with individual differences in prior achievement, in order to attain comparable levels of subsequent achievement. The reader is referred to Smith (1977) for an introductory overview of mastery learning. Let me note in passing that a recent study found "time to learn" to be a better predictor of academic achievement than IQ in the fourth, fifth, and sixth grades (Gettinger and White 1979).

Over the past 20 years, alternative methods of schooling have been derived from theories and principles of developmental psychology. In particular, Piaget's theory, which highlights commonalities rather than differences among children's intelligence, has been the source for numerous early childhood and high school science programs. (See De Lisi 1979 for an introduction to Piaget's theory.)

From the perspective of those who would like to see IQ testing done away with, it is unfortunate that many of these alternative models for educational instruction and evaluation have not been attempted on a large scale. Time, effort, and money are needed for implementation and for assessment of their effectiveness. If there is a public call for alternatives to IQ testing, then there must be an attendant public commitment to this search for alternative ways to meet the needs of individual pupils in schools. If ways can be found to minimize or even capitalize upon individual differences in ability to learn in school, we will have gone a long way toward identification of an intelligent alternative to IQ testing.

Bereiter, C. IQ and elitism. Interchange, 1976-1977, 7, no. 3, 34-42.

Bronowski, J. The ascent of man. Boston: Little, Brown, & Co., 1973.

Brown, A. L., and French, L. A. The zone of potential development: Implications for intelligence testing in the year 2000. Intelligence, 1979, 3, 255-273.

De Lisi, R. The educational implications of Piaget's theory and assessment techniques. ERIC TM Report 68, ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, Princeton, N.J., 1979.

Furth, H. G. Piaget, IQ, and the nature-nurture controversy. Human Development, 1973, 16, 61-73.

Gage, N. L., and Berliner, D.C. Educational psychology. 2d ed. Chicago: Rand McNally, 1979.

Gettinger, M., and White, M.A. Which is the stronger correlate of school learning? Time to learn or measured intelligence? Journal of Educational Psychology, 1979, 71, no. 4, 405-412.

Glaser, R. The process of intelligence and education. In The nature of intelligence, ed. L.B. Resnick. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

---. Adaptive education: Individual diversity and learning. New York: Holt, Rinehart and Winston, 1977.

Goodnow, J.J. Concepts of intelligence and its development. In Studies in cross-cultural psychology, ed. N. Warren. London: Pergamon, in press.

Horn, J. L. Trends in the measurement of intelligence. Intelligence, 1979, 3, 229-240.

Klein, S.D. Psychological testing of children. A consumer's guide. The Exceptional Parent Press, 1977.

Neisser, U. General, academic, and artificial intelligence. In The nature of intelligence, ed. L.B. Resnick. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Resnick, L.B. Changing conceptions of intelligence. In The nature of intelligence, ed. L.B. Resnick. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

---. The future of IQ testing in education. Intelligence, 1979, 3, 241-253.

Scarr, S. From evolution to Larry P., or what shall we do about IQ tests? Intelligence, 1978, 2, 325-342.

Smith, J.K. Perspectives on mastery learning and mastery testing. ERIC TM Report 63, ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, Princeton, N.J., 1977.

---. The role of measurement in the process of instruction. ERIC TM Report 70, ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, Princeton, N.J., 1979.

Terman, L.M. Binet's method of measuring intelligence. In Readings in child psychology, ed. W. Dennis. 2d ed. Englewood Cliffs, N.J.: Prentice-Hall, 1963.

Tyler, L.E. The intelligence we test. An evolving concept. In The nature of intelligence, ed. L.B. Resnick. Hillsdale, N.J.: Lawrence Erlbaum Associates; 1976.

APPENDIX 1

The list of tests as arranged by Binet in 1911

Age 3

1. Points to nose, eyes, and mouth.
2. Repeats 2 digits.
3. Enumerates objects in a picture.
4. Gives family name.
5. Repeats a sentence of 6 syllables.

Age 4

1. Gives his or her sex.
2. Names key, knife, and penny.
3. Repeats 3 digits.
4. Compares 2 lines.

Age 5

1. Compares 2 weights.
2. Copies a square.
3. Repeats a sentence of 10 syllables.
4. Counts 4 pennies.
5. Unites the halves of a divided rectangle.

Age 6

1. Distinguishes between morning and afternoon.
2. Defines familiar words in terms of use.
3. Copies a diamond.
4. Counts 13 pennies.
5. Distinguishes pictures of ugly and pretty faces.

Age 7

1. Shows right hand and left ear.
2. Describes a picture.
3. Executes 3 commissions, given simultaneously.
4. Counts the value of 6 sous, 3 of which are double.
5. Names 4 cardinal colors.

Age 8

1. Compares 2 objects from memory.
2. Counts from 20 to 0.
3. Notes omissions from pictures.
4. Gives day and date.
5. Repeats 5 digits.

Age 9

1. Gives change from 20 sous.
2. Defines familiar words in terms superior to use.
3. Recognizes all the pieces of money.
4. Names the months of the year, in order.
5. Answers easy "comprehension questions."

Age 10

1. Arranges 5 blocks in order of weight.
2. Copies drawings from memory.
3. Criticizes absurd statements.
4. Answers difficult "comprehension questions."
5. Uses 3 given words in not more than 2 sentences.

Age 12

1. Resists suggestion.
2. Composes 1 sentence containing 3 given words.
3. Names 60 words in 3 minutes.
4. Defines certain abstract words.
5. Discovers the sense of a disarranged sentence.

Age 15

1. Repeats 7 digits.
2. Finds 3 rhymes for a given word.
3. Repeats a sentence of 26 syllables.
4. Interprets pictures.
5. Interprets given facts.

Adult

1. Solves the paper-cutting test.
2. Rearranges a triangle in imagination.
3. Gives differences between pairs of abstract terms.
4. Gives 3 differences between a president and a king.
5. Gives the main thought of a selection that he or she has heard/read.

[1] This was taken from Terman (1963).

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ERIC/TM Report 74

INTELLIGENCE, INTELLIGENCE TESTING, AND SCHOOL PRACTICES

by

Richard De Lisi

Rutgers University

This paper is intended for parents and teachers who want to know more about intelligence testing as a schooling practice. In defining intelligence, social scientists have agreed that it is partly a function of the values of the given culture; that a distinction should be made between "the ability" and "manifestations of that ability"; and that intelligence can be validly measured.

The psychometric approach, the behavioristic approach, and the developmental approach to the study of intelligence are described. The history of school testing practices is traced, and Binet's work is discussed. Types of intelligence tests are described, and the properties of each type are discussed. The interpretation of IQ test scores is explained, including the issues of test validity and test bias. Finally, alternations to IQ testing are suggested.

- -

ORDER FORM

Please send ___ copies of ERIC/TM Report 74, "Intelligence, Intelligence Testing, and School Practices," at $4.50 per copy.

Name _____

Address _____ Zip _____

Total enclosed $_____

Return this form to:

ERIC/TM

Educational Testing Service

Princeton, NJ 08541

TM 800 779

RECENT TITLES

IN THE ERIC/TM REPORT SERIES

#73 - Measuring Attitudes Toward Reading, by Ira Epstein. 120p. $9.50

#72 - Methods of Identifying Gifted Minority Students, by Ernest M. Bernal. 32p. $4.50

#71 - Sex Bias in Testing: An Annotated Bibliography, by Barbara Hunt. 85p. $5.00.

#70 - The Role of Measurement in the Process of Instruction, by Jeffrey K. Smith. 29p. $3.50.

#69 - Client Assessment: A Manual for Employment and Training Agencies, by Thomas E. Backer.

Volume I: Introduction to Assessment Program Development. 58p. $4.00.

Volume II: Assessment Techniques. 49p. $4.00. ($7.00 for both volumes.)

#68 - The Educational Implications of Piaget's Theory and Assessment Techniques, by Richard De Lisi. 37p. $5.00.

#67 - Does the Rasch Model Really Work? A Discussion for Practitioners, by R. Robert Rentz and Charlotte C. Rentz. 23p. $3.00.

A complete ERIC/TM publications list can be obtained from ERIC/TM, Educational Testing Service, Princeton, NJ 08541.